ML HW10 Report

物理三 潘阜承 B08202036

1. **根據你最好的實驗結果，簡述你是如何產生transferable noises,Judge**

**Boi上Accuracy降到多少?**

In my best settings, I used DIM-MIFGSM as my main method to attack the given model. In DIM-MIFGSM, I first defined a function to randomly resize the input image within the range [32\*0.9, 32]. Then, I padded zeros around the image to the size of 32\*32 in a random manner. Finally, I set the probability of transformation to 0.5. After defining the transformation function, I used it when inputting adversarial input to the model in order to generate the loss. Then, the procedure is basically the same as MIFGSM. With this setting, my accuracy on JudgeBoi has reduced to 0.090.

1. **當source model為resnet110\_cifar10(from Pytorchcv), 使用最原始的fgsm**

**攻擊在dog2.png的圖片。**

1. **請問被攻擊後的預測的class是錯誤的嗎？**

Ans: No, the predicted class is still correct with 96.35% confidence.

1. **實作jpeg compression (compression rate=70%) 前處理圖片, 請問**

**prediction class是錯誤的嗎？**

Ans: No, prediction class is still correct with the confidence increasing to 98.85%.

1. **Jpeg compression為什麼可以抵擋adversarial attack, 讓模型維持高正確率？**

Ans: b.